USING ICPSR RESOURCES TO TEACH SOCIOLOGY*

The focus on quantitative literacy has been increasingly outside the realm of mathematics. The social sciences are well suited to including quantitative elements throughout the curriculum but doing so can mean challenges in preparation and presentation of material for instructors and increased anxiety for students. This paper describes tools and resources available through the Inter-university Consortium for Political and Social Research (ICPSR) that will aid students and instructors engaging in quantitative literacy across the curriculum. The Online Learning Center is a source of empirical activities aimed at undergraduates in lower-division substantive courses and Exploring Data through Research Literature presents an alternative to traditional research methods assignments. Searching and browsing tools, archive structures, and extended online-analysis tools make it easier for students in upper-division undergraduate and graduate courses to engage in exercises that increase quantitative literacy, and paper competitions reward them for doing so.

LYNETTE F. HOELTER
The University of Michigan

FELICIA B. LECLERE
The University of Michigan

AMY M. PIENTA
The University of Michigan

RACHAEL E. BARLOW
Trinity College

JAMES W. MCNALLY
The University of Michigan

OVER THE LAST TWO DECADES, colleges and universities have taken on the challenge of teaching quantitative literacy to students in all majors (Hunt 2004; Madison 2006; Paulos 2001; Wallman 1993). As the quantitative literacy (QL) or "numeracy" movement expands from mathematics into other disciplines, some have suggested that the social sciences are particularly well suited to teaching students quantitative skills. Students can engage in solving a problem which makes data meaningful and interesting (Shrimplin and Yu 2004; Sweet and Strand 2006; Steen 2007). More recently, through the Integrating Data Analysis pro-

ject, the American Sociological Association has examined the typical sociology curriculum and sounded a similar call by challenging sociology departments to introduce students to "data analysis early, frequently, and sequentially throughout the curriculum" (Howery and Rodriguez 2006:23). This echoes other curricular initiatives that suggest that students generally learn more when relevant examples are offered that begin with a foundation and build on skills and previous knowledge, particularly when there are numerous opportunities for working with similar skills ("practicing") and when supportive help is available. These curricular innovations can be found in the information literacy movement and writing across the curriculum efforts (Hunt 2004).

Quantitative or statistical literacy has been defined in various ways, but perhaps the most succinct definition was presented by Wallman in her presidential address to the

---

*The main content of this paper was presented as a Teaching Workshop at the American Sociological Association meetings in New York, NY, August 2007. Please direct all correspondence to Lynette Hoelter, ICPSR, 426 Thompson Street, Ann Arbor, MI 48106-1248; email: lhoelter@umich.edu.

American Statistical Association (1993): "Statistical literacy is the ability to understand and critically evaluate statistical results that permeate our daily lives—coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions" (p. 1). This definition captures not only the essence of quantitative literacy but also the important role it plays in daily life. Schield (2004) further adds to the definition by distinguishing between information, statistical, and data literacy. That is, *information literacy* is concerned with finding and critically evaluating information to answer questions, *statistical literacy* involves the use of statistics as evidence for arguments, and *data literacy* focuses on the need to understand and work with data. Taken together, these three types of literacy embody the skills central to sociology as a discipline and, therefore, are critical to undergraduate and graduate curricula.

However, Howery and Rodriguez (2006), Madison (2006) and others note the difficulties of including quantitative literacy in undergraduate classes. For example, instructors must often re-work syllabi and assignments to do so (Grauerholz and Gibson 2006), and while collaborative work is a useful strategy for building QL, it is often met with student resistance (Caulfield and Persell 2006). Additionally, adequate practice of these new skills is required, often using current news or research information. This removes some of the instructor's control over what can and will be covered as the type of information available for such exercises in any given semester is unpredictable (Madison 2006). Additionally, faculty and students themselves sometimes have difficulty separating quantitative literacy from traditional mathematics as they evoke similar anxieties and stresses. On the other hand, because quantitative literacy is different than general math, even math faculty cannot simply fall back on traditional approaches when teaching quantitative literacy (Howery and Rodriguez 2006, Madison 2006).

This paper recognizes the importance of teaching QL as well as the difficulties of implementing strategies to accomplish that goal. We discuss several ways in which the Inter-university Consortium for Political and Social Research (ICPSR) and its resources can enhance students' quantitative literacy in coursework across the undergraduate and graduate curricula. It should be noted that the two are not meant to be mutually exclusive – advanced undergraduate and lower-level graduate courses can successfully use these resources and share similar assignments with minor modifications in grading and expectations. Finally, we discuss several general tools available through the ICPSR website that enhance teaching and research at all levels.

GOALS, RESOURCES, AND TOOLS

Including Data in Introductory-level Courses

One of the largest challenges facing instructors of lower-division undergraduate classes is introducing students to topics that showcase the discipline while capturing their attention and demonstrating relevance to students' lives. These are challenging enough without instructors having to include QL at the same time. Working with real data to demonstrate examples of class concepts, however, not only engages students in more active learning but also builds a foundation that integrates research methodology and statistics into the content of the discipline. It also decreases the disconnect between substantive topics and methods and statistics introduced later in undergraduate careers (e.g., Caulfield and Persell 2006; Howery and Rodriguez 2006). The social sciences often attract students who are uncomfortable with math and statistics, but using data-based examples in early courses may work to attract students with some quantitative skills and interests to the discipline.

ICPSR has prioritized making data holdings more accessible to faculty and students and is developing a resource that will make

integrating data analysis into introductory-level classes easier for instructors. Several key themes emerged in focus groups and in-depth interviews about such topics with teaching faculty. Instructors wanted the ability to 1) quickly locate relevant data that are easy to work with and work to demonstrate the concept(s) and 2) customize any materials to their own teaching approach and syllabus. The result is the Online Learning Center (OLC: www.icpsr.umich.edu/OLC), a centralized site focused on bringing data into the classroom. The materials included in the OLC are: 1) teaching/learning modules that identify relevant, usable datasets quickly and use them to demonstrate a core social science concept, 2) accompanying search and data tools, 3) a tool that allows instructors to set up a space for their class to download data, share resources, and work within the ICPSR context ("MyClass"), and 4) links to other instructional materials both within and outside of ICPSR. Future development includes a depository where faculty can share activities in the same form as the current learning modules and a mechanism for feedback on various activities.

The core of the site is made up of data-driven learning guides (DDLGs)—activities that instructors can pull out for student use either through demonstration in the classroom or by having students access the website directly and work through the activities. These guides match major sociology and methods/statistics topics to ICPSR datasets and demonstrate the concepts through basic data analysis using the online Survey Documentation and Analysis (SDA) software. Concepts chosen for inclusion are common across textbooks typically used for Introduction to Sociology courses. Multiple activities on particular topics also make them useful for more narrowly focused substantive courses such as courses on the family or stratification. Students do not need to learn, nor do institutions have to purchase licenses for, any statistical software to be able to complete the exercises—everything is accessible online. To alleviate any appre-hension about technology that might still accompany basic computer data analysis (e.g., Howery and Rodriguez 2006), the exercises have clickable links so that students are able to get results without even engaging the SDA interface. Each DDLG is set up like a lesson plan, including a goal statement, or learning objective of the activity; a brief discussion of the concept and how it is used in sociology; a short description of the dataset—accompanied by a link to the full dataset, should the instructor or student want to do further analyses on his/her own; an analytic application that addresses some question(s) related to the concept; notes about interpreting the results; and a bibliography of additional resources or data on the topic.

The DDLGs make use of "real" data from the ICPSR holdings—for example, anything from the National Health Interview Surveys to the American National Election Surveys, Survey of Russian Marriages, or the ABC Opinion Polls might be used for a particular guide. The analyses are tested and simplified by collapsing categories on variables with many values (if using cross-tabs or some other statistical test that requires categorical data) and defining "non answers" as missing so as to make visual presentation and interpretation easier for students. Students are also given assistance in interpreting results for the focal questions of the exercise but additional links allow for flexibility in further examining other variables and basic research questions. This resource will get students working with data in the first classes in the discipline and strengthen both QL skills and the understanding that the field of sociology involves the use of research and evidence to study social phenomena.

A New Approach to Teaching Research Methods

Instructors of upper-level undergraduate research methods classes often include projects in which students analyze data to answer a research question of the students' choosing. The goal is to have the students

emulate what scholars do by producing a "mini" version of a peer-reviewed journal article. The hope is that through emulation students will be able to learn the important steps in the research process and develop an understanding of the core issues in sociological research. But some problems can arise with such a project. Often, students have difficulty identifying the large range of research questions that they could apply to a specified dataset. Furthermore, these students struggle to operationalize their hypotheses using the variables available in the dataset and so have difficulty articulating how the dataset speaks to a given research question. Barlow, building on her experience working in a library, argued that students asked to conduct the emulation project are not typically privy to the iterative, non-linear process researchers undertake in developing a research project and the resulting manuscript. Students asked to read an article or articles and create their own project attempt to emulate the research process but are forced to do so only by examining the one or two isolated products of that process (i.e., a few journal articles) they have seen in other classes. This lack of exposure to the scholarly process *outside* of some isolated journal articles is why, according to Barlow, students may have a problem coming up with good research questions that lead to the analysis of quantitative data.

Barlow suggests that a more useful way of engaging students in the research process is to use a model in which the students analyze not data in the traditional sense but the *objects* of social science discourse: the scholars, datasets, and publications that compose a network of scholarly exchange. Barlow's hope is that by asking students to explore the relationships between scholars, their publications, and the datasets from which they publish, it will force students to be self-reflective of their discipline, avoid (albeit temporarily) merely emulating the practices of professional social scientists, and instead analyze and learn from those practices.

To encourage instructors to try this alternative method to teaching methods, Barlow has developed a tool called *Exploring Data through Research Literature* (EDRL: www.icpsr.umich.edu/EDRL/). EDRL is a series of proposed exercises that instructors can ask students to complete. EDRL exploits ICPSR's Bibliography of Data-Related Literature, a searchable database of 41,000 citations for published and unpublished works using data in ICPSR's collection. The Bibliography allows users to enter the system in multiple ways: 1) by searching on a *dataset* which allows students to see the many scholars who have used the same dataset to produce multiple papers, 2) by searching on an *author*, showing how one scholar may use multiple datasets to answer related questions, or, conversely, might use a single dataset to produce multiple papers, and 3) directly from an *article* which would take students to the dataset used to answer the question of interest to them.

EDRL's exercises all follow a similar structure: students begin with an "entry article" that the instructor has carefully selected for both its content (e.g., interesting hypotheses, inclusion of important methodological terms, or significant methodological procedures) and for the nature of its inclusion in ICPSR's Bibliography (whether it appears in the Bibliography and to how many other datasets/articles/authors is it connected). The exercises all ask the students to move from reading the entry article and answering questions about it to examining the description of the data associated with it and to using the Bibliography to discover other articles that examine the same data or attempt to address the same research question. In each exercise, students are ultimately asked to use the Bibliography to select their own "exit article," based on specific criteria related to how the exit article relates to the entry article with which they began the activity. For example, one exercise asks the student to find an exit article written by the same author as the entry article and that analyzes the same dataset. Another exercise asks the student to find an

exit article that uses the same dataset but is written by a different author (see Table 1). Students can then compare exit and entry articles to determine how authors use theory, methodology, and data to answer their respective research questions (Barlow N.d.).

The foundation of the EDRL is that important information can be learned about social science methodology by comparing the works of researchers who employ that methodology and that the Bibliography of Data-Related Literature is both "a means to an end and an end in itself. Students can use networks of data and literature to find articles suitable for comparison. As they navigate these networks, they will also begin to uncover a bird's-eye view of their discipline" (Barlow N.d.).

The Use of Secondary Data in Graduate Courses

Usage statistics show that graduate students represent the largest single group of individuals who download ICPSR data and that their usage is increasing over recent years (e.g., 10,175 downloads in Spring 2005 compared to 34,023 downloads in Spring 2007). It is likely that at least some of these downloads are due to class assignments that require data analysis as faculty members often require graduate students to use data for assignments and papers. Students also use datasets as the basis for professional presentations or publications as well as master's theses and dissertations. Our focus here is on the ways in which data can be used in the classroom for either replication of previous work or original analyses.

Asking students to replicate an article of interest in sociology or other fields is a valuable way to socialize students to the expectations and practices of the discipline, teaches them the norms of scientific transparency, and helps them get published early in their careers (King 2006). Direct replication typically begins with a topic in which the student is interested and involves finding an article in a sociological journal, acquiring the data used in the article, and replicating the analyses done by the initial author. The next steps include determining precisely if the results were replicated, building on the replicated results to try to improve them, and doing some controlled methodological experiments to try to advance knowledge on the topic (King 2006).

Two ICPSR tools make this assignment easier for students by removing the uncertainty that the data of the original author can be acquired. The Publication-Related Archive (www.icpsr.umich.edu/ICPSR/pra/index.html) gives students access to data deposited to ICPSR by authors. These data are not reprocessed by ICPSR; rather, the Publication-Related Archive serves as a "holding place" expressly for the purpose of making those data available for replication of an article or book. Specialized computer programs or recode statements, a file describing the materials in "sufficient technical and substantive detail to permit replication of findings," and any other relevant documentation received and contact information for the investigator are presented along with the data (ICPSR N.d.). The second tool is the Bibliography of Data-Related Literature (http://www.icpsr.umich.edu/ICP

Table 1. Description of Exercises for *Exploring Data Through Research Literature* Tool

	Relative to Entry Article, Exit Article is...			
	By Same Author	Uses Same Dataset	By New Author	Uses New Dataset
Exercise 1: Other Works	X	X		
Exercise 2: Other People		X	X	
Exercise 3: Other Datasets			X	X

SR/citations/index.html), described briefly above. Each citation in this database is linked to the dataset on which that article is based, making the task of tracking down the data as easy as clicking a link and downloading the data in the preferred format. Both tools are searchable by dataset name, author name, subject words, or general topic. Citations in the databases also link to full-text versions of the articles if they are available at the user's campus location.

When students are tasked with doing or proposing original analyses, using the search tools and the organization of the ICPSR website often makes the initial task of finding relevant work or data less overwhelming. In addition to the Publication-Related Archive and the Bibliography of Data-Related Literature, searching and browsing the collection guarantees that students will be able to easily link data and literature on topics of interest. The basic search tool located on the ICPSR homepage works much like any library search tool, finding keywords in dataset and publication titles or descriptions. Browsing the collection is a good strategy for those who have some general interests but are not yet sure what keywords to use to narrow the focus. Users can find data organized by topical archive, thematic categories, or series, as well as datasets that can be used with the online analysis system, data recently added or updated, or unpublished census data. Such distinctions can help point students in the right direction by steering them toward a collection of items that are best suited to their needs.

ICPSR's topical archives are often a good starting point for graduate students with interests in areas such as demography, minorities, criminal justice, child care and early education, substance abuse and mental health, or aging. For example, the National Archive of Computerized Data on Aging (NACDA) has been used in graduate courses to help students explore the health and well-being of the elderly as well as their roles in family and social relationships throughout life. The NACDA Archive preserves and distributes almost 2,000 unique studies that address issues of aging and health and performs three central functions (with support from NIA) by 1) acquiring and preserving data sets of scientific importance to the gerontological research community, 2) distributing data and documents in a manner that makes their use easy and cost effective for research, and 3) contributing to the intellectual vitality of the gerontological sciences. NACDA also contains specially created datasets such as an analytic file of elderly households from all 52 parts of the 2000 Census five-percent public use microdata files. Searches with the same functionality as those described above are available within the archive so that students who are interested in an aging-related topic are able to quickly and easily locate relevant data and publications. Exercises for students using the archive can include anything from quick online analyses to using the literature and data tools for writing a proposal in the format of a grant application.

General Tools for Making Data Instruction and Analysis Easier

Each of the resources discussed above represents a utility that addresses some of the barriers instructors and students face in including quantitative literacy in college classrooms. The resources presented in this section are more general and help support QL in a less-structured way as they are not part of specific assignments but tools to make assignments easier or initiatives to reward students for learning QL skills. We specifically discuss the online analysis tool, tutorials and teaching aids, and new ICPSR initiatives for undergraduate instruction beyond those previously noted.

Several online tools and applications make the data analysis easier for many studies. The Survey Documentation and Analysis (SDA) interface, developed at UC-Berkeley and mentioned above, is available with studies not only in the Aging archive or the Online Learning Center, but throughout the ICPSR data holdings. Currently

there are more than 370 studies that can be used with SDA and more are added constantly. SDA provides both simple descriptive statistics like frequencies and means as well as multivariate analyses including logistic and linear regression. Quick Tables is another tool that offers cross-tabulations on previously selected combinations of variables. Selected studies, especially those that are a part of the Minority Data Resource Center, also have three additional tools including a sample exploration tool to allow students to see if there are enough cases to answer their question about a particular subgroup of the population, a subset extractor so that users can choose to download only data for individuals who fit various demographic characteristics—by age or race categories for example—and a statistical software syntax tool that translates recode commands given to the SDA system into syntax for the major statistical packages like SPSS and SAS (www.ICPSR.umich.edu/MDRC). Additionally, a variable search tool makes comparisons of different measurement strategies across datasets straightforward. These online tools offer several advantages over traditional statistical software, particularly for students new to quantitative analysis. The advantages are: 1) data are prepackaged with documentation and labeled so results of each analysis are well marked and more easily interpretable, 2) online tools are available without any software beyond a web-browser, and 3) tutorials and instructions are available online which provide easy instructions for use.

Tutorials and teaching aids can be found throughout the ICSPR website. For example, the Data User Tutorial (www.icpsr.umich.edu/ICPSR/help/newuser.html) provides both a general introduction to using data as well as information specific to using ICPSR's site such as a list of acronymns, search tips, and the like. One of the archives has also developed a detailed tutorial for using SDA for students and users who are new to data analysis and might find the program daunting (www.icpsr.umich.edu/SAMHDA/tutorial). Lastly, the MyClass tool is a new initiative that allows instructors to provide classes a time-limited mass registration for ICPSR so that they can share data and activities with all students in a class (www.icpsr.umich.edu/cgi-bin/bob/myclass).

Finally, two initiatives are aimed at continuing to foster interest and competency in data analysis in undergraduates. First, students who successfully complete a research project using ICPSR data can be rewarded through the Undergraduate Paper Competition (both for general ICPSR data and, new for 2007-2008, a separate competition for papers using data from the Minority Data Resource Center: www.icpsr.umich.edu/ICPSR/prize/). Second, students who have had a taste of data analysis and want to learn more about social science research can apply for one of a number of undergraduate summer internships at ICPSR (www.icpsr.umich.edu/ICPSR/careers/internship.html).

CONCLUSION

Teaching students to be comfortable thinking about, evaluating, and using data as evidence in arguments is one of the tasks sociologists take to heart. Quantitative literacy is a natural fit in courses like Research Methods or Statistics, but a greater challenge lies in introducing QL skills early in the curriculum as a part of more substantively-oriented courses, when students are less likely to be "expecting" it. Demonstrating throughout the undergraduate curriculum the ways in which sociologists rely on data of all types to draw conclusions about social phenomena is a large task. ICPSR has developed a number of tools that can assist instructors in enhancing QL across the curriculum. The graduate curriculum often fits naturally into a QL model, but resources like the Online Learning Center and Exploring Data through Research Literature encourage faculty to engage undergraduates in this type of critical thinking and analysis from the beginning of their college careers. Online analysis tools also make bringing data into the classroom for demonstration or exploration less onerous

as instructors can spend more time on the content of the lessons rather than spending valuable class time teaching more complicated statistical packages before students need to learn them. The online environment is also one in which students are generally comfortable, perhaps removing some of the "technology anxiety" typically associated with data analysis for some students (Howery and Rodriguez 2006). The resources discussed here enable instructors to give students at all levels a taste of what sociologists do and how they do it.

REFERENCES

Barlow, Rachael. (N.d.) Exploring Data through Research Literature. Retrieved September 21, 2007 (http://www.icpsr.umich.edu/EDRL).

Caulfield, Susan and Caroline Hodges Persell. 2006. "Teaching Social Science Reasoning and Quantitative Literacy: The Role of Collaborative Groups." *Teaching Sociology* 34(1):39-53.

Grauerholz, Liz and Greg Gibson. 2006. "Articulation of Goals and Means in Sociology Courses: What We Can Learn from Syllabi." *Teaching Sociology* 34(1):5-22.

Howery, Carla and Havidan Rodriguez. 2006. "Integrating Data Analysis: Working with Sociology Departments to Address the Quantitative Literacy Gap." *Teaching Sociology* 34(1):23-38.

Hunt, Karen. 2004. "The Challenges of Integrating Data Literacy into the Curriculum in an Undergraduate Institution." *IASSIST Quarterly* Summer/Fall:12-5.

Inter-university Consortium for Political and Social Research (N.d.) Publication-Related Archive Description. Retrieved September 28, 2007 (http://www.icpsr.umich.edu/ICPSR/pra/index.html).

King, Gary. 2006. "Publication, Publication." *Political Science & Politics* 39(1):119-25.

Madison, Bernard L. 2006. Pedagogical Challenges of Quantitative Literacy. Retrieved September 20, 2007 (http://www.statlit.org/PD (http://www.statlit.org/PDF/2006MadisonASA.pdf).

Paulos, John Allen. 2001. *Innumeracy: Mathematical Illiteracy and Its Consequences*. New York: Hill and Wang.

Scheitle, Chris. 2006. "Incorporating Web-Based Data Analysis into the Introductory Curriculum." *Teaching Sociology* 34(1):80-6.

Schield, Milo. 2004. "Information Literacy, Statistical Literacy, and Data Literacy." *IASSIST Quarterly*, Summer/Fall:6-11.

Shrimplin, Aaron K. and Jen-chien Yu. 2004. Focusing in on Student Learning Outcomes: How SDA Helped Us Get Data into the Classroom. *IASSIST Quarterly* Summer/Fall: 55-7.

Steen, Lynn Arthur. (N.d.) "Quantitative Literacy: Why Numeracy Matters for Schools and Colleges." The Mathematical Association of America. Retrieved September 5, 2007 (http://www.maa.org/features/QL.html).

Sweet, Stephen and Jerry Strand. 2006. "Cultivating Quantitative Literacy: The Role of Sociology." *Teaching Sociology* 34(1):1-4.

Wallman, Katherine K. 1993. "Enhancing Statistical Literacy: Enriching Our Society." *Journal of the American Statistical Association* 88(421):1-8.

Lynette Hoelter, PhD, is a research investigator at ICPSR involved in projects focused on using data in the classroom, among other duties. She is also a lecturer in the sociology department and the Survey Methodology Program at the University of Michigan. Her research interests include the relationship between social change and marital quality, gender in families, and the study of family and relationship dynamics more broadly.

Felicia LeClere, PhD, is the director of the ICPSR project entitled Data Sharing for Demographic Research, which is a new project funded by NICHD to assist data producers in making data ready for public release. In addition to her responsibility at DSDR, LeClere manages the Minority Data Resource Center and is cross-appointed as associate research scientist in the Population Studies Center at ISR. Her research interests include the health of immigrants and the influence of social and physical geography on health disparities.

Rachael Barlow is the Social Science Data Coordinator at Trinity College in Hartford, Connecticut, where she teaches data-related topics across the college curriculum and serves as the Official Representative for ICPSR. She is also completing her doctorate in sociology at Indiana University in Bloomington. In her dissertation, she examines how librarians and library buildings co-construct librarian professional identities.

Amy Pienta, PhD, is Acquisitions Director at ICPSR. She is also a research affiliate of the University of Michigan Center on the Demography of Aging and the Population Studies Center at the University of Michigan. At ICPSR, she oversees new data acquisitions as well as identifying new data collections in the social sciences and strategic planning for new data acquisitions. Pienta's primary research interests are

retirement and health in later life.

James McNally, PhD, is the Director of the NACDA Program on Aging at the University of Michigan. A gerontologist and demographer, Dr. McNally was trained at Georgetown University and Brown University and specializes in family support of the aged with a focus on disability. McNally directs an aging program that supports the National Archive of Computerized Data on Aging, a data repository containing over 1,600 unique studies on aging and health. Dr. McNally lectures widely both in the United States and internationally on issues related to the aging life-course.